Tellus (2004), 56A, 89–101
Printed in UK. All rights reserved

TELLUS

Tentative probabilistic temperature scenarios for northern Europe

By RASMUS E. BENESTAD*, *The Norwegian Meteorological Institute, PO Box 43 Blindern, 0313, Oslo, Norway*

(Manuscript received 31 October 2002; in final form 13 June 2003)

ABSTRACT

An ensemble of global climate scenarios from different coupled atmosphere–ocean general circulation models are evaluated against re-analysed observations in terms of temperature trends over the period 1958–1998. The spatial warming rate patterns for the "past" derived from the individual models exhibit large differences, but the median values of the ensemble appears to capture some of the spatial structure observed. A probabilistic approach describing the past warming rates is evaluated, and the combination of large inter-model differences and the good evaluation results justifies a probabilistic view on the future climate.

Regional temperature scenarios are presented for northern Europe in the form of probability distributions, based on spatially interpolated empirically downscaled trends, derived using a multi-model ensemble as well as various downscaling options. The estimation of probabilities is based on the assumption that there is no common systematic error in the various model scenarios, the IS92a scenario is valid for the future and their scatter portrays the actual uncertainty. The results point to a model consensus of stronger warming over land during winter, with maximum warming over Finland. The same analysis for July suggests an overall weaker summertime warming with a smaller difference between the coasts and the interior.

1. Introduction

Increasing atmospheric concentrations of so-called greenhouse gases, due to burning of fossil fuels, give a reason for major concern because this kind of increase is expected to lead to an enhanced greenhouse effect (Houghton et al. 2001). In order to study how our climate might respond to the changes in the associated radiative forcing, numerical global climate models (GCMs) have been constructed in order to simulate how the climate may be affected. Although the GCM results roughly agree regarding the global mean temperature, different GCMs often give diverging warming estimates on regional scales such as the north European and the Arctic region (Räisänen, 2001a,b; Benestad et al. 2002b). The discrepancy amongst the GCMs may be due to several factors such as individual GCM bias, different integration strategies, internal variability and different spin-up histories. Although some GCMs are more realistic in terms of reproducing the present-day climate, sophistication, and spatial resolution than others, it is difficult to pick the most probable regional climate scenario from a range of simulations. Some of the most advanced GCMs (e.g. HADCM3 and NCAR-CSM) that do not employ flux-correction tend to produce a cold bias in the description of the "present-day" climate in the Greenland–Iceland–Norwegian Sea region (Holland et al. 2001; Benestad et al. 2002a). The range of GCM estimates, however, may contain useful information about the GCMs in general since the scatter in the climatic trend estimates may be taken as a crude measure of GCM uncertainties (Räisänen and Palmer, 2001). It is important to keep in mind that whereas a large scatter implies large uncertainties, a small scatter does not mean that the uncertainty is small because it is possible that all the GCMs are wrong.

The use of multi-model ensembles may improve the GCM-based data sample, both in terms of more independent realizations and in canceling systematic biases specific to an individual GCM. Such ensembles have been used to study how increased greenhouse gas concentrations may alter the range of temperature and precipitation (sometimes referred to as "super ensembles") (Houghton et al. 2001; Räisänen, 2002; Räisänen and Alexandersson 2003). The improved sampling also provides better estimates of probability distribution functions (PDFs), allowing for a "probabilistic approach" to climate projections. In their pioneering work, Palmer et al. (2000) derived probability estimates from multi-model ensemble integrations and used them as input to a decision model assisting users of seasonal forecasts. This work was later extended to future climate projections (Räisänen and Palmer, 2001; Palmer and Räisänen, 2002).

*e-mail: rasmus.benestad@met.no

Table 1. Details on the different downscaled estimates. The Table gives a summary of the different simulations: 2 (ECHAM4 GSA & GSDIO) + 1 (NCAR-CSM) + 3 (CCCMA) + 1 (CSIRO) + 2 (ECHAM3) + 1 (GFDL) + 1 (HADCM3) + 4 (HADCM2) + 1 (NCAR-DOE) + 1 (CCSR/NIES) = 17. The first column n gives the number of trend estimates for each model used for constructing the empirical probability distributions, i.e. sums up number of ticks in columns 5–8. The second column lists the GCM, and column 3 gives the emission scenario: 'GSA' is the IPCC IS92a, 'b006' denotes a 1% pa annual increase, whereas 'GSDIO' is similar to 'GSA' but with the additional effects of tropospheric ozone and aerosols' indirect effect on clouds. Columns 5–8 indicate which regions were used for the predictors in the downscaling, where the natl region is 90°W–40°E, 40°N–75°N, neurope 30°W–40°E, 50°N–70°N, *nordic* 20°W–40°E, 50°N–75°N, and scan 0°W–30°E, 55°N–75°N)

n	Model (member)	Run	Predictor	*natl*	*neurope*	*nordic*	*scan*	Resolution (long × lat)
4	ECHAM4/OPYC3	GSDIO	T(2 m)	v	v	v	v	T42/L19 ∼2.8 × 2.8
3	ECHAM4/OPYC3	GSDIO	Φ(700–500)		v	v	v	T42/L19 ∼2.8 × 2.8
3	ECHAM4/OPYC3	GSDIO	T(850 hPa)	v	v		v	T42/L19 ∼2.8 × 2.8
2	ECHAM4/OPYC3	GSA	T(2 m)			v	v	T42/L19 ∼2.8 × 2.8
4	NCAR-CSM	b006	T(2 m)	v	v	v	v	∼2.8 × 2.8
6	CCCMA 1-3	GSA	T(2 m)			v	v	T32/L10 ∼3.8 × 3.7
2	CSIRO	GSA	T(2 m)			v	v	R21/L9 ∼5.6 × 3.1
6	ECHAM3/LSG 1,2	GSA	T(2 m)		v	v	v	T21 ∼5.6 × 5.5
2	GFDL	GSA	T(2 m)			v	v	R15 ∼7.5 × 4.4
4	HADCM3	GSA	T(2 m)	v	v	v	v	L19 3.75 × 2.5
8	HADCM2 1-4	GSA	T(2 m)			v	v	L19 3.75 × 2.5
2	NCAR-DOE	GSA	T(2 m)			v	v	∼7.5 × 4.4
2	CCSR/NIES	GSA	T(2 m)			v	v	∼5.6 × 5.5
48	**sum**							

In order to assist decision making and make useful scenarios for the future, it is necessary to assess the skill of the GCMs. There has been a number of evaluation studies of GCMs (Houghton et al. 2001). Much of the past evaluation has primarily involved global mean surface temperatures (Tett et al. 1999; Stott et al. 2001; Forest et al. 2002). When GCMs are used to study the response of regional or local climates to an enhanced greenhouse effect, it is desirable to evaluate their ability to reproduce the regional and local climatic trends of the past, although only a few studies have focussed on this issue.

The spatial resolution of the most sophisticated global GCMs is typically around 3 × 3 deg^2 (Table 1) and important features for local climates, such as mountain ranges and valleys, are overly smoothed (Benestad, 2002b). The Scandinavian climate is strongly influenced by the local geography, which comprises mountain ranges, fjords, valleys, forests, with different characteristic climates. The present state-of-the-art GCMs are incapable of giving a realistic description of the local climate in regions with a complex physiography (Houghton et al. 2001), partly because they have too coarse a spatial resolution to adequately represent the important geographical features. Another issue is that GCMs are believed to have a lower limit on their skillful spatial scale (Grotch and MacCracken, 1991; von Storch et al. 1993a).

Local climate scenarios can nevertheless be derived using empirical downscaling techniques where geographical information (implicitly) as well as the relationship between large-scale climatic anomalies and the local climate (explicitly) are taken into account (e.g. Zorita and von Storch, 1997, 1999). It is also possible to "nest" higher-resolution models of a limited region and take the results from the global climate models as boundary conditions (dynamical downscaling) to improve the description of the regional climate (Christensen et al. 1998, 2001; Kidson and Thompson 1998; Haugen et al. 1999; Murphy, 1999, 2000). In this study, however, large-scale climatic patterns are used to derive information concerning the local climates through the means of empirical downscaling. Empirical downscaling can involve a number of different methods from simple regression between local climatic elements and climatic indices to advanced multi-variate techniques and neural nets (von Storch et al. 1993b; Heyen et al. 1996; Zorita and von Storch, 1997; Corte-Real et al. 1998; Kidson and Thompson, 1998; Kilsby et al. 1998; Schubert, 1998; Wilby et al. 1998; Deliang and Hellström 1999; Zorita and von Storch, 1999; Huth and Kyselý, 2000; Busuioc et al. 2001; Hellström et al. 2001; Beckmann and Buishand, 2002; Huth, 2002; Oshima et al. 2002), and there has been a considerable development in the methodology over recent years. Table 2 gives a summary of part of this development and a selection of references to multi-variate methods relevant to empirical downscaling. Some of the most recent work has entailed an improvement over earlier work, both in terms of downscaling methodology as well as improvements in the GCMs. This study uses the downscaled results from Benestad (2002b) and takes the analysis a step further by including additional information (e.g. constructing maps of probabilities and making use of geographical information) not utilized in the previous analyses. A combination of a geographical model and residual kriging ensures more realistic maps than the kriging-only analysis in Benestad (2002b),

Table 2. Examples of various methods used in multi-variate analysis and empirical downscaling listed together with a selection of relevant references. (This list is not comprehensive.)

Empirical orthogonal functions (EOFs)	Lorenz (1956), North et al. (1982)
Canonical correlation analysis (CCA)	Glahn (1968), Anderson (1958)
Kriging	Matheron (1963)
Combination of principal component analysis (PCA) and CCA	Barnett and Preisendorfer (1987) Preisendorfer (1988)
Empirical/statistical downscaling	Beckmann and Buishand (2002), Hellström et al. (2001) Heyen et al. (1996), Kilsby et al. (1998), Kidson and Thompson (1998) Deliang and Hellström (1999), von Storch et al. (1993b) Zorita and von Storch (1999), Huth (2002), Huth and Kyselý (2000) Oshima et al. (2002), Zorita and von Storch (1997)
Use of multi-model ensembles for analysing probabilities	Räisänen and Alexandersson (2003), Räisänen (2002) Räisänen and Palmer (2001)
Common PCA and EOFs	Flury (1988), Sengupta and Boyle (1993), Sengupta and Boyle (1998) Barnett (1999)
Mixed EOFs (combined PCA)	Kutzbach (1967), Bretherton et al. (1992)
Use of CCA in empirical downscaling	von Storch et al. (1993a), Heyen et al. (1996) Huth (2002)
Common EOFs (cEOF) in empirical downscaling	Benestad (2001)
Mixed-common EOFs (mcEOF) in empirical downscaling	Benestad et al. (2002b)
Empirical downscaling of multi-model ensembles	Benestad (2002a)
Kriging and downscaling (EOF space)	Biau et al. (1999)
Kriging to produce spatial maps from downscaling	e.g. Benestad (2002b)

and probabilistic estimates are less sensitive to outliers than ensemble mean values. The geographical modelling is also used to examine dependences on parameters such as the distance from the coast, altitude, latitude and longitude. Furthermore, reconstructions of trends and probabilities by individual GCMs are evaluated against observations.

In order to avoid confusion, the term "model" will henceforth be used when referring to a statistical model either used in empirical downscaling or describing the geographical dependence, whereas "GCM" is used when referring to global climate models. However, the term "multi-model ensemble" will refer to GCMs. A distinction will also be made between the results from simulations with GCMs ("global scenarios") and results derived through empirical downscaling ("downscaled scenarios"). A trend is in this paper defined as a best-fitting linear time dependence of a time-series, while a "time-series" may represent single measurements from a climate station as well as chronological series of gridded realizations from a GCM simulation or gridded analyses.

The following section describes the data used in the analysis and gives a summary of the methods used for deriving the results. The methods include the evaluation of past temperature trends in the North Atlantic region, the probabilistic approach, and using geographical dependences to produce maps of warming trends. Since the details of the empirical downscaling method are published elsewhere, only a brief summary is given here on this aspect. Section 2 is followed by the results of the evaluation of the multi-model ensemble in terms of past trends and probabilistic maps. The paper concludes with a discussion and a conclusion.

2. Data and methods

2.1. Data

The data against which the GCMs were evaluated consisted of monthly means taken from the National Center for Environmental Prediction (NCEP) reanalysis (Kalnay et al. 1996). The NCEP reanalysis data does not give a "perfect" description of the real world, and is also subject to errors and uncertainties. However, the 2-m temperature used in this study gives a good reproduction of the temperature anomalies seen in climate station records in Scandinavia and northern Europe (Benestad, 1999, 2002a; Benestad et al. 2002b) and is therefore believed to be appropriate for model evaluation. The GCM simulations included 16 global climate scenarios following the IS92a emission scenario (Experiments "GSA" in Table 1 were retrieved from the Intergovernmental Panel on Climate Change (IPCC) Internet site[1]) as well as results from an integration with the NCAR-CSM GCM (taken from NCAR's Internet site[2]). The NCAR-CSM results were

[1] http://ipcc-ddc.cru.uea.ac.uk/dkrz/dkrz_index.html
[2] http://www.cgd.ucar.edu/csm/experiments/b006/downloadable.html

from a 1% yr^{-1} increasing CO_2 experiment from present-day level (355 ppm) (experiment "b006" in Table 1). A total of 17 transient integrations were produced by 10 different GCMs (Table 1), of which the HADCM2 model has produced four, ECHAM3 two and the CCCMA three ensemble members by being integrated with different initial conditions. The ECHAM4 GCM was integrated using two different set-ups: with ("GSDIO") and without ("GSA") indirect aerosol effects.

The predictands comprised 85 homogenized North Atlantic Climatological Dataset (NACD) (Frich et al. 1996) station monthly mean temperature series from northern Europe and 30 Norwegian station series from the Norwegian Meteorological Institutes climatological archive.

2.1.1. The downscaled scenarios. The downscaled temperature trends were taken from Benestad (2002b), who appended time series of the gridded monthly mean 2-m temperature (T(2 m)) anomalies from the GCMs to the gridded reconstructions of observed monthly mean anomalous T(2 m) from the Norwegian Meteorological Institute (Benestad, 2000). The GCM results were bi-linearly interpolated onto the same grid as the observations (i.e. 5 × 5 deg^2 longitude–latitude resolution) before merging the data sets. Figure 1 illustrates this pre-processing of data. The combined data set was subject to an empirical orthogonal function (EOF) analysis. EOFs of combined data sources are known as "common EOFs" (cEOF) (Barnett 1999; Benestad 2001) in the climate research community, more widely known as common principal component analysis (Sengupta and Boyle, 1993, 1998; Flury, 1988). The cEOFs describe modes with identical structures in the observations and the GCM results, and are associated with combined time series that describe their temporal variations for both the observations and the GCM data. The part of the time series representing the observations is used for the calibration of the empirical downscaling models, whereas the part describing the GCM results is used for the projection of climatic trends.

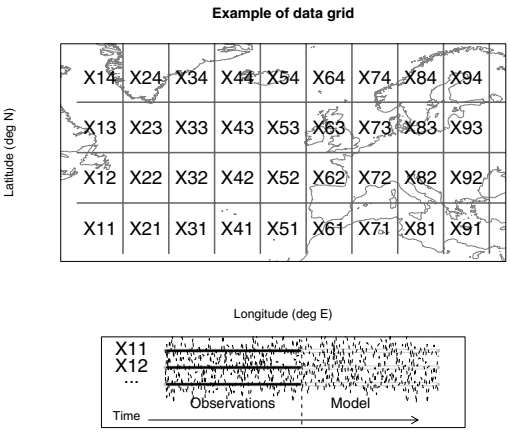

Fig 1. A schematic illustration of the pre-processing required for the common EOF analysis. First single monthly means (e.g. January) are extracted from the data, and the mean values are subtracted. The extracted model data are then interpolated onto the same grid as the observations. Each grid-box in the data sets contains a time series (vector) (e.g. X11 in the upper part of the figure). The model data (e.g. $X11_{model}$) is combined with the observations ($X11_{obs}$) so that the time series for the model results (white section of the line in the lower part) is attached to the end of the time series representing the observations (black section of the line in the lower part) for each grid-box. Common EOF analysis involves an ordinary EOF analysis applied to the combined data set. The part of the principal components representing the observations are used for model calibration and the section describing the GCM provides the input for downscaling.

The downscaling model calibration was based on a step-wise screening of the 20 leading modes in conjunction with empirical models based on canonical correlation analysis (CCA). The step-wise screening employed a cross-validation analysis (Wilks, 1995) based on the Pearson correlation. The downscaling analysis was applied to predictor data of different types and of various regional coverage (Table 1) as there is no *a priori* information defining the optimal predictor variable or domain. The predictors were chosen from four different regions referred to as *natl*, *neurope*, *nordic* and *scan*. For each location the analysis gave 48 different trend estimates for January, April and October, respectively, and 46 estimates for July. The July scenarios included fewer estimates because the smallest domain (spanning 0°W–30°E/55°N–75°N) was associated with low skill for downscaling models based on ECHAM4-GSDIO 700–500 hPa thickness and T(850 hPa). The cEOF method and the empirical downscaling are described and evaluated in detail by Benestad (2001, 2002a,b).

2.2. Methods

2.2.1. Evaluation of GCMs. In order to carry out a direct inter-comparison, the individual GCM results were bi-linearly interpolated onto the same grid as the NCEP reanalysis (i.e. 2.5 × 2.5 deg^2 longitude–latitude resolution). The evaluation of the climate scenarios was based on comparison between the 'past' linear temperature trend at each respective grid-box with corresponding 1958–1998 trends derived from the NCEP reanalysis. Trends were estimated through the means of a linear regression against time over the interval "1980"–"2050" for the scenarios and "1958"–"1998" for the 'past'[3]. The spatial correlation and root-mean-square error (rmse) were calculated for the spatial trend maps over the region 90°W–60°E, 0°N–88°N for the January, April, July and October months, respectively, for the individual GCMs as well as ensemble mean and median (NCAR-CSM was excluded from the model evaluation since this integration only started in "1990", but the NCAR-CSM were included in the projections for the future).

[3]Internal variations do not correspond between the model results and the real world, and the initial conditions of a GCM simulation has an impact on the subsequent regional temperature evolution. A distinction will therefore be made between the actual time and the model time stamp by expressing the latter in apostrophes.

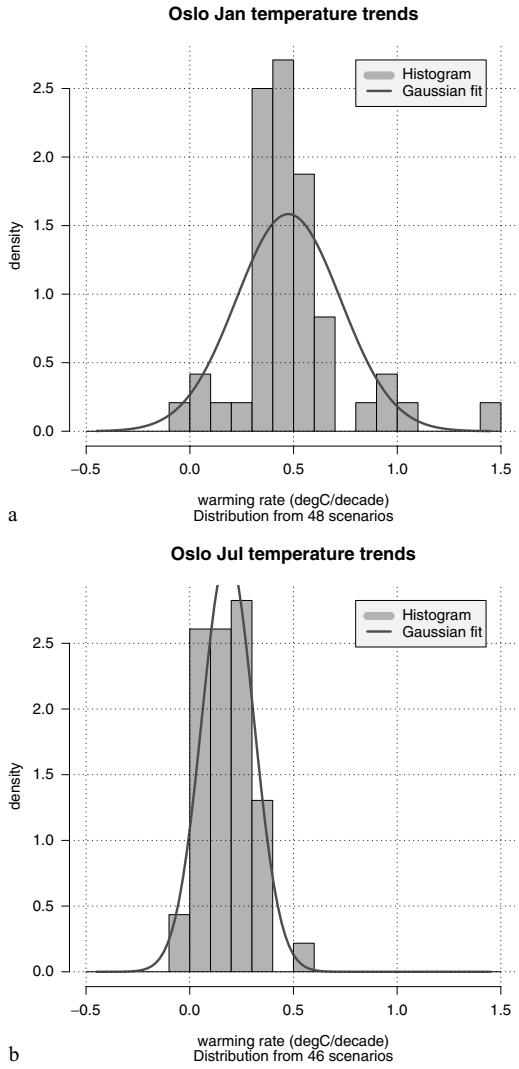

Fig 2. Histogram showing the distribution of the linear best-fit to downscaled warming trends in Oslo for January (a) and July (b). These downscaled scenarios were derived by Benestad (2002b) using the monthly mean temperature from the NACD set as predictand as well as the predictors and climate simulations listed in Table 2.

The present work is an extension of earlier work documented in Benestad (2002a) by presenting scenarios in terms of probabilities similar to those in Räisänen and Palmer (2001), and a minimum requirement for using a probabilistic approach must be that it gives a reliable description of the past trends. (This criterion also applies to deterministic descriptions of the trends.). These probabilities are plausible for the future, provided that the IS92a scenario is considered representative of the future and the multi-model ensemble spread gives an unbiased estimate of the probability density distribution. The probabilistic approach is illustrated by Fig. 2, which shows histograms of 1980–2050 warming trends for Oslo in January (a) and July (b). These histograms can be used to make an estimate of the probability of the warming rate exceeding a given value, and the probability estimate (p_i) for grid-box i ($i = 1, \ldots, N$) was estimated by counting the number of ensemble members yielding a past warming rate (y_i) exceeding a threshold value (y_*) and then dividing by the total number of ensemble members. In this example, the predicted warming is strongest in January and the histogram peaks at 0.5 °C decade^{-1}. There is also a long tail at the upper end of the scale. For July, the various GCM estimates have a smaller spread, with the majority of the GCMs projecting a trend of 0.15 °C decade^{-1} for the period 1980–2050.

The Brier score (Wilks, 1995, p. 259) was used for the evaluation of probabilities (1958–1998 period). The Brier score is estimated according to

$$\mathrm{BS} = \frac{1}{N}\sum_{i}^{N}[p_i - \mathcal{H}(y_i - y_*)]^2 \tag{1}$$

where $\mathcal{H}$ is the Heaviside function:

$$\mathcal{H}(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0. \end{cases} \tag{2}$$

It is important to keep in mind that these probability estimates do not necessarily reflect the "true probability" of actual future warming exceeding given threshold rates, but are merely a measure of how well the various GCMs agree. These "probabilities" are also conditional on the IS92a emission scenario as well as the assumption that the GCM spread gives a good representation of the true uncertainty (i.e. no common GCM shortcomings), and they will henceforth be referred to as "conditional probabilities".

2.2.2. The refinement of the downscaled scenarios. Benestad (2002b) used kriging analysis (Matheron 1963) in order to construct spatial maps describing how future warming (i.e. the multi-model ensemble mean) may vary geographically. Although this kind of spatial interpolation gives an approximately realistic representation, it does not take into account the fact that local warming rates may vary with the distance from the coast, altitude, latitude and longitude. In order to produce more realistic maps, this geographical information must be included in the spatial analysis. The results presented here are "refined" further from those of Benestad (2002b) by taking into consideration geographical information not utilized in the earlier work, in addition to examining conditional probability estimates instead of multi-model ensemble mean values. The conditional probabilities (fraction of ensemble members) of the warming rate exceeding a given threshold value (chosen to be 0.50 °C decade^{-1} for January and 0.25 °C decade^{-1} for April, July and October; see the above discussion and Fig. 2) derived for each site was used in a geographical model based on a multiple regression analysis against distance from the coast, altitude, latitude and longitude. The geographical models for January, April, July and October are summarized in Table 4 in Section 3.2, which

shows the dependences on the distance from the coast, altitude, latitude and longitude as well as the uncertainties associated with these. The model details in Table 4 are listed in the form of coefficient estimates, standard error estimates, *t*-values and probabilities (*p*-value) of null-hypothesis (zero coefficient) being true. Also shown are the estimates for the R^2 (variance explained), the *F*-statistic ("strength" of the regression), the degrees of freedom and the *p*-value for the entire regression. It is important to keep in mind that the geographical models derived here may not be valid for other parts of the world. A kriging analysis similar to Benestad (2002b) was used for the spatial interpolation of the residuals from the regression analysis.

3. Results

3.1. Model evaluation

Figure 3a shows the geographical distribution of linear trends in January derived from the NCEP re-analysis, based on a linear regression against time. This analysis represents the observed warming trends over 1958–1998, and the results suggest strongest warming in the vicinity of the ice-edge. One explanation for this rapid warming may be a thinning of the sea-ice (Rigor et al. 2002) or a retreat in the sea-ice (Benestad et al. 2002a). Figure 3b shows the corresponding analysis based on

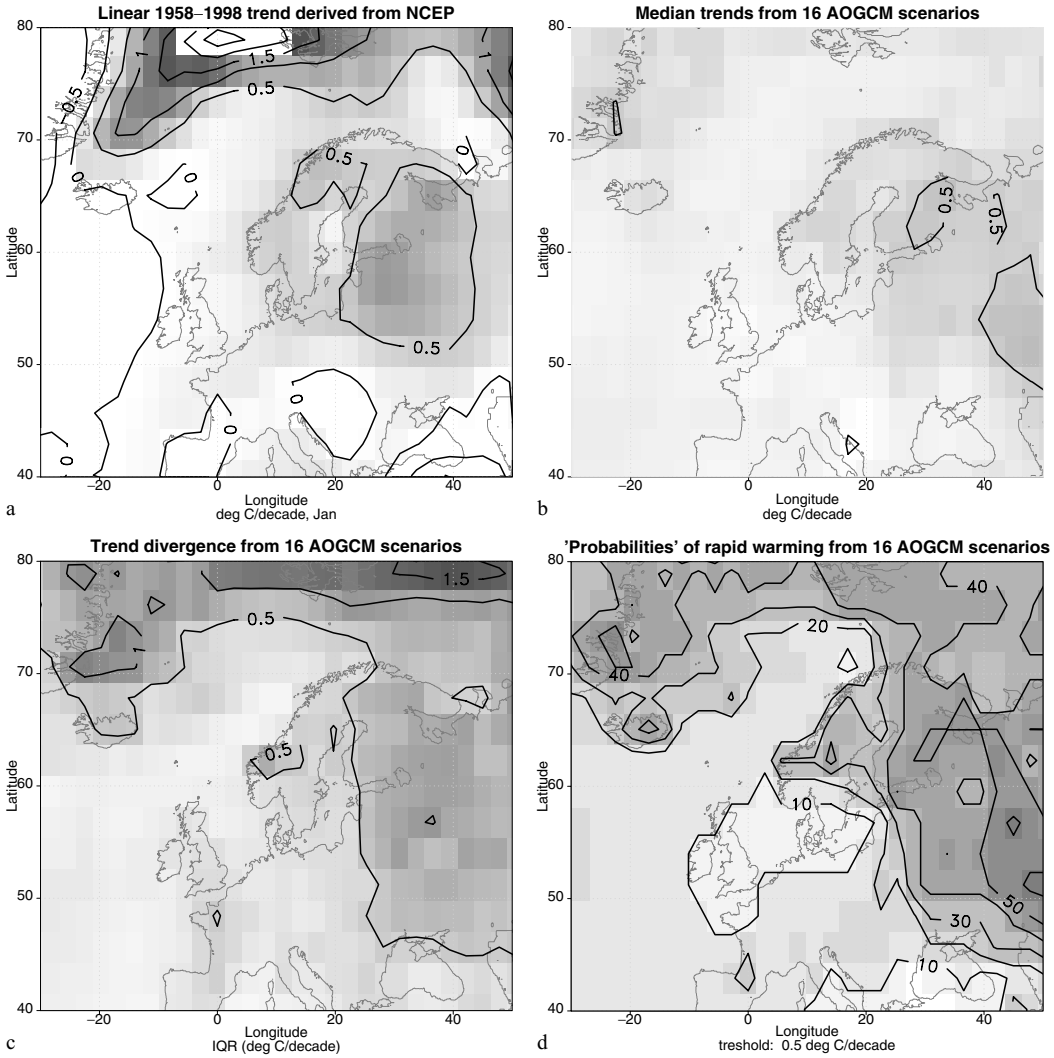

Fig 3. Comparison between observed (a) and the multi-model ensemble median simulated linear trends (b) corresponding to January 1958–1998. Panel (c) shows the inter-quantile range (IQR) of the multi-model ensemble trend estimates, and panel (d) shows the conditional probabilities of the mean 1958–1998 January warming rates exceeding 0.5 °C decade^{-1}.

Table 3. Spatial pattern correlation (cor) and root-mean-square error (rmse) over 90°W–60°E, 0°N–88°N between 1958–1998 linear trends from NCEP reanalysis T(2 m) and from GCM T(2 m). The results are expressed as "cor/rmse" for each of the months January, April, July and October. Also shown is the Brier score for the probabilistic scenarios. High correlation but low rmse and Brier scores (associated with a threshold rate of 0.50 °C decade^{-1} for January and 0.25 ° decade^{-1} for April, July and October) signify good skill. The analysis is made for 1672 grid boxes.

GCM run	January	April	July	October
ECHAM4 GSDIO	0.49/0.63	0.51/0.36	0.05/0.32	0.46/0.42
ECHAM4 GSA	0.30/0.65	0.17/0.41	0.14/0.31	0.37/0.45
ECHAM3 1	−0.62/0.82	−0.01/0.42	0.02/0.30	−0.29/0.48
ECHAM3 2	0.44/0.52	0.19/0.38	−0.01/0.30	−0.18/0.42
CCCMA 1	0.11/0.71	0.00/0.45	−0.13/0.36	−0.02/0.43
CCCMA 2	−0.05/0.74	0.24/0.39	−0.08/0.35	0.06/0.55
CCCMA 3	0.12/0.67	0.26/0.40	−0.01/0.34	0.03/0.49
CSIRO	0.30/0.69	0.37/0.38	−0.19/0.33	−0.39/0.48
GFDL	−0.05/0.73	0.00/0.45	0.08/0.33	0.30/0.58
HADCM3	−0.60/1.13	−0.26/0.64	−0.01/0.35	0.14/0.49
HADCM2 1	−0.12/0.68	−0.35/0.67	−0.02/0.33	−0.07/0.42
HADCM2 2	0.55/0.57	0.55/0.43	0.06/0.29	0.31/0.53
HADCM2 3	0.01/0.76	0.06/0.54	−0.02/0.32	−0.22/0.53
HADCM2 4	0.65/0.72	0.36/0.55	0.04/0.33	0.16/0.47
NCAR-DOE	−0.59/5.18	−0.27/4.08	0.12/2.79	−0.13/4.34
CCSR/MIES	0.49/4.56	0.14/3.70	−0.16/2.82	0.30/5.03
Ensemble mean	−0.04/0.85	−0.04/0.52	−0.01/0.44	0.15/0.70
Ensemble median	0.21/0.62	0.32/0.37	0.03/0.28	0.20/0.40
Brier score	0.12	0.06	0.16	0.22

the multi-model ensemble median derived from the climate scenarios. The GCMs underestimate the warming over the eastern Russia peninsula and fail to capture the retreat of the sea-ice in the Greenland and Barents seas. Figure 3c shows the dispersion within the ensemble, which suggests that there is a high degree of discrepancy amongst the GCMs in the vicinity of the ice edge. The conditional probability analysis shown in Fig. 3d is based on multi-model ensemble distributions similar to those shown in Fig. 2.

Table 3 shows the skill scores from the evaluation of the individual GCMs as well as the ensemble. The NCAR-DOE and NIGS results are clearly unrealistic: they have root-mean-square errors that are almost an order of magnitude greater than those of the other models. No GCM stands clearly out as superior, although the ECHAM4-GSDIO and two of the HADCM2 members score high in three out of the 4 months examined. April gives on average the best correlation scores and the best Brier score, however, it is important to keep in mind that the January Brier score is not directly comparable to the others as the January threshold value is different because the warming is much stronger in January than in the other months. The July month, on the other hand, tends to give lower correlation scores on average than the other months[4], suggesting that the summer scenarios are more difficult than the rest of the year. October gives the worst Brier scores. However, there are large differences amongst the various ensemble members based on the same GCM (ECHAM3, CCCMA and HADCM2), suggesting that model evaluation based on merely 41 yr in the past (the length of the NCEP data) is difficult. The large spread may be a result of a weak trend, the short series used in the analysis, and hence the difficulty in defining a trend (i.e. a low signal-to-noise ratio).

The scores associated with the multi-model ensemble mean are poor, due to the sensitivity to clearly unrealistic GCMs (NCAR-DOE and NIGS) included in the ensemble. The ensemble median, on the other hand, is not so sensitive to outliers, and systematically yields better scores. Thus the GCM evaluation based on the 1958–1998 period (Table 3) suggests that it is better to use the multi-model ensemble median rather than the ensemble mean. Probabilistic results also have a lower sensitivity to the unrealistic GCMs than the ensemble mean. The large inter-GCM spread in combination with the good (i.e., low) Brier scores in Table 3 seems to justify a probabilistic approach.

Figure 4 shows a comparison between the multi-model ensemble median of the January trends estimated for the period 1980–2050 (a) and corresponding conditional probabilities of the warming rate exceeding 0.5 °C decade^{-1} (b). It is clear from this figure that the probabilistic approach gives slightly different information than the ensemble median (e.g. less than 20% of the GCMs suggest a January warming rate over the British Isles exceeding 0.5 °C decade^{-1} and 60–70% of the GCMs indicate warming rates greater than 0.5 °C decade^{-1} over parts of Russia), although the large-scale structure is similar.

3.2. *Downscaled climate scenarios*

Figure 5 shows the geographical distribution of the conditional probabilities of a warming rate exceeding 0.5 °C decade^{-1} in January, based on distributions of trends estimated from downscaled scenarios similar to that for Oslo in Fig. 2. The main difference between Figs. 5 and 4 is that the former takes into account local geographical effects through the means of downscaling and by utilizing the dependences of the conditional probabilities associated with downscaled trend-estimates to the distance from the coast, altitude, latitude and longitude. This map indicates that the likelihood for a future warming rate exceeding 0.5 °C decade^{-1} is low near the coasts, assuming the IS92a scenario. It is furthermore likely that the winter warming will be stronger than 0.5 °C decade^{-1} over interior Finland and southern Sweden, as the conditional probability estimates are above 70% in these areas. A comparison between Figs. 4b and 5 shows that that there is a good agreement between the large-scale structures of the scenarios derived directly from the global climate scenar-

[4]The rmse score was not used in the comparison between seasons since the trend magnitudes are expected to be seasonally dependent.

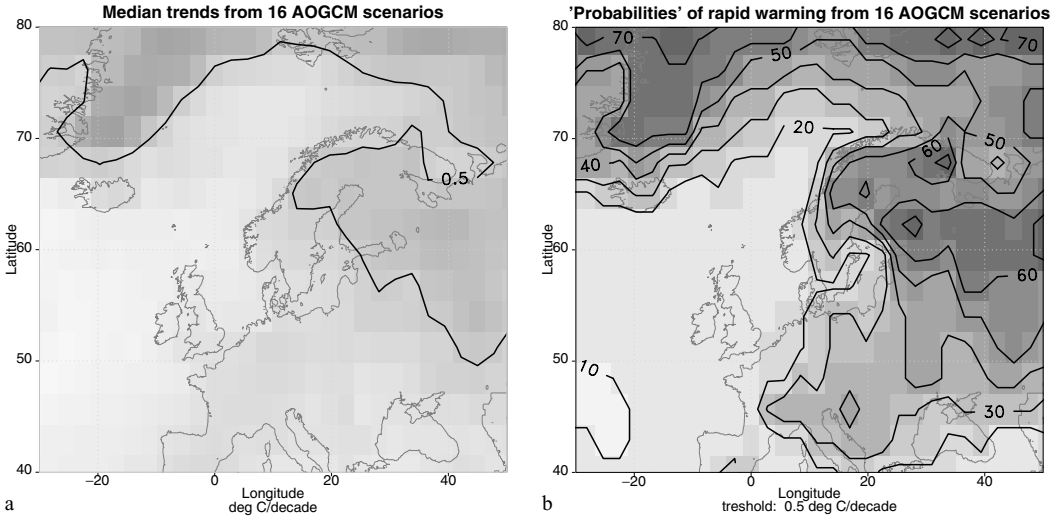

Fig 4. A comparison between the GCM ensemble median linear trend for January 1980–2050 (a) and conditional probability of a warming greater than 0.5 °C decade^{-1} (b).

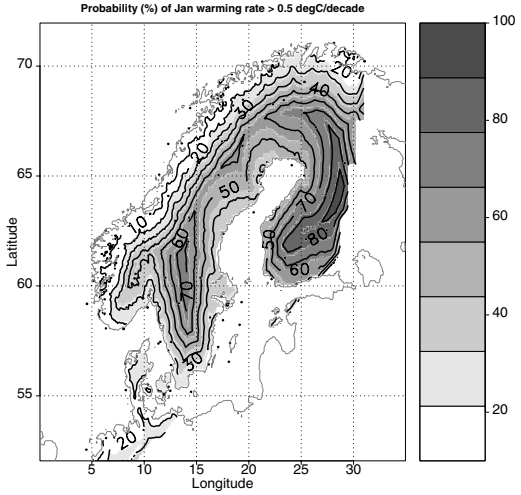

Fig 5. Probability estimates of the warming exceeding a threshold value of 0.5 °C decade^{-1} in January for "1980"–"2050", derived from downscaling of the multi-model ensemble and based on the assumption that there is no common systematic bias amongst the various GCMs. The map was produced using the multiple geographical regression model (Table 4) and kriging of the residuals from the geographical model. The stations included in the analysis are marked as black dots.

ios and the downscaling, but it is also apparent that the local details are enhanced by downscaling. The conditional probabilities are in general higher in the downscaling analysis (over land). Figure 5 reveals new features not seen in the analysis of the multi-ensemble median (Fig. 4a) warming trends or the conditional probabilities derived directly from the GCM results (Fig. 4b). There is a larger proportion of downscaled scenarios suggesting a rapid warming over southern Sweden than over northern Sweden, and this information cannot be seen in a map of multi-model ensemble median (Fig. 4a) or ensemble means (Benestad 2002b, Fig. 9a). Hence, the probability can give more information than just the ensemble median or mean. The ANOVA results in Table 4 suggest that the important geographical parameters in January are the distance from the coast and longitude. The geographical model accounts for 42% of the observed spatial structure, and thus a substantial part of the spatial variations of the conditional probabilities in winter must be incorporated through kriging.

Figure 6 shows corresponding empirically downscaled scenarios for April, but for threshold values of 0.25 °C decade^{-1}. The analysis points to low conditional probabilities ($p < 15\%$) for warming rates exceeding 0.25 °C decade^{-1} in southwestern Norway, and that there are no regions with high values. The highest conditional probabilities ($p \sim 55\%$) are seen over interior Finland. In April, both latitude and longitude are important for the conditional probabilities, but the distance from the coast has a weak effect on the geographical distribution (Table 4). An R^2-value of 61% suggests that conditional probabilities have a stronger dependence on geographical features in spring than in the other seasons.

The warming is also much weaker in July than in January. Figure 7 shows a similar map of conditional probability estimates for the warming rate exceeding 0.25 °C decade^{-1} in July. The highest conditional probabilities are found over southern Finland and the interior southern Scandinavia. The downscaled results exhibit a clear latitude–longitude dependence (Table 4). Most (68%) of the spatial variations in the conditional probabilities in the summer, however, are not related to the distance from the coast, altitude, latitude and longitude.

In October the highest conditional probabilities are found in the interior parts of southern Sweden and Norway and near the coast of southwestern Finland. Since the threshold value used for the October analysis is 0.25 °C decade^{-1}, the results are still

Table 4. The analysis of variance (ANOVA) of the geographical multiple regression model, $y = c_0 + c_0$ dist $+ c_0$ alt $+ c_0$ lat $+ c_0$ lon, for January, April, July and October. The independent variable "dist" is the distance from the coast, "alt" is the altitude, "lat" is the latitude, and "lon" is the longitude of the location of the downscaled scenarios. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

January				
	Estimate	Std. error	t-value	Pr(>|t|)
(Intercept)	0.394 001	20.805 723	0.019	0.985
dist	17.358 068	2.998 599	5.789	6.80e−08***
alt	−0.005 374	0.006 237	−0.862	0.391
lat	0.174 924	0.342 191	0.511	0.610
lon	0.583 544	0.126 104	4.627	1.02e−05***
$R^2 = 0.4239$	F-statistic = 20.23	on 4 and 110 DF	p-value = 1.635e−12	
April				
(Intercept)	−59.271 664	13.387 953	−4.427	2.26e−05***
dist	3.839 196	1.929 522	1.990	0.0491*
alt	−0.003 259	0.004 013	−0.812	0.4186
lat	1.218 826	0.220 191	5.535	2.13e−07***
lon	0.769 375	0.081 145	9.481	6.66e−16***
$R^2 = 0.611$	F-statistic = 12.45	on 4 and 110 DF	p-value = 0	
July				
(Intercept)	6.730e+01	1.140e+01	5.902	4.04e−08***
dist	4.061e+00	1.643e+00	2.471	0.015*
alt	−7.806e−05	3.418e−03	−0.023	0.982
lat	−8.371e−01	1.875e−01	−4.464	1.96e−05***
lon	3.328e−01	6.911e−02	4.816	4.71e−06***
$R^2 = 0.316$	F-statistic = 12.71	on 4 and 110 DF	p-value = 1.557e−08	
October				
(Intercept)	51.052 393	19.427 661	2.628	0.009 82**
dist	2.808 711	2.799 987	1.003	0.31801
alt	0.023 623	0.005 824	4.056	9.35e−05***
lat	−0.491 098	0.319 526	−1.537	0.12718
lon	0.674 214	0.117 752	5.726	9.05e−08***
$R^2 = 0.3873$	F-statistic = 17.38	on 4 and 110 DF	p-value = 4.45e−11	

consistent with the strongest warming in January (Fig. 8). The geographical analysis suggests a systematic variation in conditional probabilities with altitude as well as longitude (Table 4). The conditional probabilities in October are not sensitive to the distance from the coast or the latitude, and 61% of the spatial variance must be added in the form of kriging.

4. Discussion

The downscaled results from the multi-model ensemble point to a large proportion of trend estimates, indicating enhanced warming over the continents in winter. The large number of scenarios, indicating a strong winter warming over Finland and southern Sweden seems to be robust and merits an explanation. The additional information seen in the local details comes from the statistical models employed in the analysis, describing the relationship between the large and small scales, although the GCMs describe much of the large-scale spatial structure. Even though the empirical models are statistical, it can be argued that they do reflect the underlying physical mechanisms explaining the patterns. The explanation for the close model agreement (80–90%) on the high warming rate in the interior of Finland does not seem to be that the January temperature in this region is relatively insensitive to variations in the North Atlantic Oscillation (NAO) ($R^2 \sim 30\%$), since there is good model agreement on the weak warming in southwestern Norway where the influence of the NAO is strong ($R^2 \sim 50\%$). Furthermore, few of the GCM simulations considered here suggest a significant trend in the NAO, and the different descriptions of the NAO can therefore be regarded as different realizations of climatic noise. The regions with highest conditional probabilities are characterized as forested with inland climates, in which during winter there would be snow on the ground and in some situations on the trees. Snow can affect the local climate through the local radiative balance, the effect on moisture and latent heat of fusion (a 'buffer' effect with respect to temperature), and it is possible that a climatic change in this region is amplified through changes in the snow cover. The cloud cover may also affect the local radiative balance, e.g. low January temperatures in Scandinavia are usually associated with clear skies. Temperature inversions associated with extremely cold conditions near the ground can take place during winter in the interior of Scandinavia, and a change in the

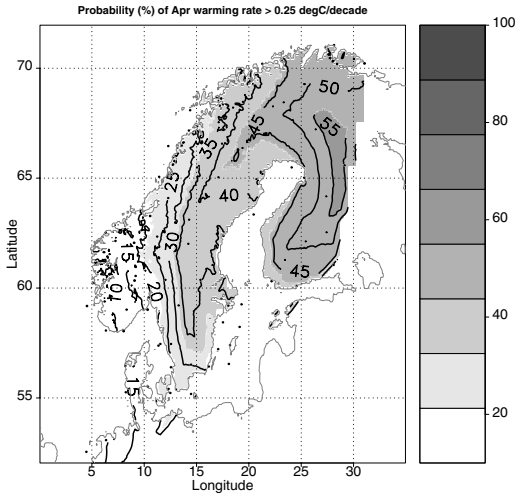

Fig 6. Same as in Fig. 5, but for the April month and a threshold of 0.25 °C decade^{-1}.

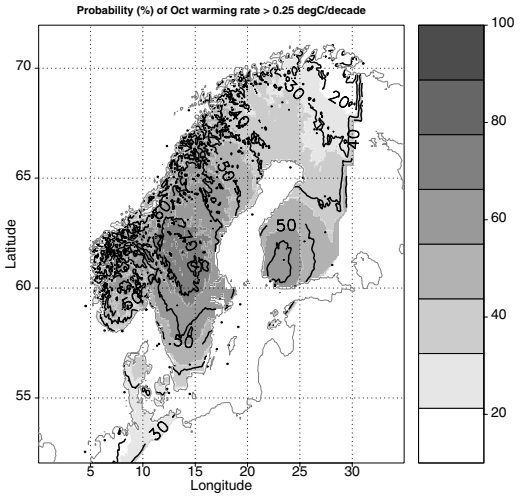

Fig 8. Same as in Fig. 5, but for the October month and a threshold of 0.25 °C decade^{-1}.

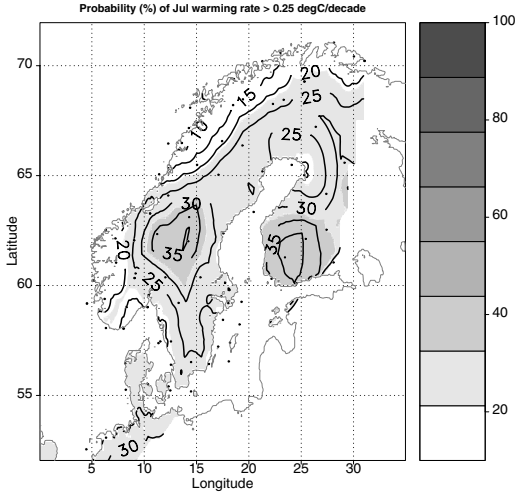

Fig 7. Same as in Fig. 5, but for the July month and a threshold of 0.25 °C decade^{-1}.

frequency of temperature inversion effects may also affect the temperature trends. The dependence of the results on the distance of the coast (Table 4) also suggests that the warming near the coasts may be moderated by the oceans. The isolated "hot spot" (four stations) over southern Sweden ($p > 60\%$) may also be a coincidence owing to the problem of multiplicity (Wilks, 1995).

A similar "hot spot" is seen in July, albeit shifted slightly to the northwest of the high conditional probabilities in January. Again, the most plausible explanation is that the enhanced warming over this part of Sweden is a result of additional geographical information from the actual world introduced by the empirical downscaling. But there is no snow in the summer so it cannot play a role in this case. A physical explanation for the geographical differences in the summer-time warming could involve humidity or cloud cover. Cloudy summer days are often cooler than clear summer days. However, it is important to keep in mind the fact that the model evaluation indicated relatively high degree of uncertainties associated with the July trends (Table 3).

The April results are associated with the best evaluation scores, and the geographical model accounts for more spatial variance than the other months examined, and the greatest conditional probabilities (∼50%) are found over eastern Finland while the lowest probabilities are seen over southern Norway. In other words, there is a model consensus on weak spring-time warming in southern Norway, and there are an increasing number of models giving faster spring-time warming towards the east and the north. One explanation for the slow warming over southern Norway could involve the influence from the sea or clouds, and an enhanced warming over the northeastern part could involve the spring-time snow cover or cloud cover.

It is only the October results that exhibit a dependence on altitude, and it is interesting to note a positive correlation between altitude and the conditional probability estimate. In other words, the conditional probability for the October warming rate exceeding 0.25 °C decade^{-1} tends to be higher in the high-altitude regions. A stronger autumn warming in the mountains could possible be related to later autumn snows in the mountains, affecting the local radiative balance, humidity and thermodynamics. Clouds may also play a role. The geographic model only can account for about 39% of the spatial probabilistic warming pattern, and there are regions not associated with mountains where a large proportion of the estimates indicate fast autumn warming.

The warming associated with the IS92a emission scenario is expected to be stronger in the future than in the past, and hence the signal-to-noise ratio is expected to be stronger for these scenarios. Moreover, as long as the emission scenarios are plausible, the multi-model ensemble will provide a plausible climate

scenario. It is nevertheless important to emphasis the fact that there are large uncertainties associated with these results. Although the GCMs roughly reproduce the geographical patterns of the past warming, they tend to underestimate the magnitude. Even if the simulation of the past did agree well with the historical observations, there is no guarantee that the future emission scenario will come true. The fact that different ensemble members of integrations with the same GCM but different initial conditions (ECHAM3, CCCMA and HADCM2) give very different evaluation scores (Table 3) suggests that internal variations are pronounced and affect the trend estimates. Allen et al. (2000) argue that the use of multi-model ensembles may be problematic because the GCMs do not necessarily span the full range of known climate system behaviour, and it is important to keep this in mind when interpreting the conditional probability estimates. The empirical downscaling may introduce additional uncertainties in the scenarios due to shortcomings in the analysis (Benestad, 2001) as linear statistical downscaling models merely provide an approximate description of the relationship between the different spatial scales. It is also possible that these statistical relationships are non-stationary in time (Wilby, 1997). It is therefore important to take into consideration not only the uncertainties associated with the GCMs but also those accompanying the downscaling. Furthermore, the trends are often not linear, and therefore a linear fit may not be the most appropriate description for the past climatic evolution (Benestad, 2003).

5. Conclusions

The evaluation of the GCM results revealed large differences between individual climate scenarios. The best evaluation scores were generally seen in April and the worst reproduction of the geographical distribution of past trends was found in October and July. The large spread in evaluation scores implies that little confidence should be attached to a single-integration scenario, and an ensemble approach is required in order to extract the climate-change signal from the internal decadal variability. The evaluation of 16 different GCM-based climate simulations in terms of the 1958–1998 warming trends also suggested that the GCM ensemble median in general underestimates the past winter-time warming rates. The rapid warming in the Greenland–Iceland Sea is absent in the GCM median, however, the GCMs do reproduce some of the observed land–sea contrasts. The enhanced warming in the Greenland–Iceland Sea is more easily seen in maps of estimated conditional probabilities. The good Brier scores associated with the probabilistic approach and the large spread in inter- and intra-GCM scores justify the use of conditional probabilities in presenting climate scenarios.

The downscaling analysis of the multi-model ensemble presented here points to high conditional probabilities regarding a rapid future warming over interior Finland and southern Sweden during winter. These results may be taken as a first estimate for a probabilistic climate forecast, assuming the IS92a emission scenario, that the systematic errors in the individual GCMs are independent, that the overall multi-model ensemble samples most of the data space of the actual climate and that the GCM ensemble is not biased.

Further work on empirical downscaling is planned which will involve a repetition of the analysis on new emission scenarios and GCM simulations following the IPCC special Report Emission Scenarios (SRES) scenarios and extend the analysis to precipitation. A new climate analysis package (clim.pact) for the R-environment[5] (Ellner, 2001; Gentleman and Ihaka, 2000) will be used in future work.

6. Acknowledgments

Ole Vignes assisted in retrieving the ECHAM4/OPYC3 GSDIO data. Ole Einar Tveito gave valuable comments about spatial interpolation and three anonymous reviewers gave valuable comments on the manuscript. This work was performed under the Norwegian Regional Climate Development under Global Warming (RegClim) program, and was supported by the Norwegian Research Council (Contract NRC-No. 120656/720) and the Norwegian Meteorological Institute. The plots in this paper were made with the GNU project's R data analysis tool and the Gebhardt (2000) kriging R-package.

References

Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. and Delworth, T. L. 2000. Quantifying the uncertainty in forecasts of anthropogenic climate changes. *Nature* **407**, 617–620.

Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis*. 1st edn. Wiley, New York.

Barnett, T. P. 1999. Comparison of near-surface air temperature variability in 11 coupled global climate models. *J. Climate* **12**, 511–518.

Barnett, T. P. and Preisendorfer, R. W. 1987. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.* **115**, 1825–1850.

Beckmann, B.-R. and Buishand, T. A. 2002. Statistical downscaling relationships for precipitation in the Netherlands and north Germany. *Int. J. Climatol.* **22**, 15–32.

Benestad, R. E. 1999. Conversion of the NCEP re-analysis data to the netCDF format and quality control. KLIMA 31/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R. E. 2000. Analysis of gridded sea level pressure and 2-meter temperature for 1873–1998 based on UEA and NCEP re-analysis II. KLIMA 03/00. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R. E. 2001. A comparison between two empirical downscaling strategies. *Int. J. Climatol.* **21**, 1645–1668. DOI 10.1002/joc.703.

Benestad, R. E. 2002a. Empirically downscaled multi-model ensemble temperature and precipitation scenarios for Norway. *J. Climate* **15**, 3008–3027.

[5]R can be downloaded freely from http://www.R-project.org/, and clim.pact can be found under the link to "contributed packages".

Benestad, R. E. 2002b. Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble. *Climate Res.* **21**, 105–125.

Benestad, R. E. 2003. What can present climate models tell us about climate change? *Climatic Change*. **59**, 311–332.

Benestad, R. E., Hanssen-Bauer, I., Skaugen, T. E. and Førland, E. J. 2002a. Associations between the sea-ice and the local climate on Svalbard. KLIMA 07/02. met. no, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R. E., Hanssen-Bauer, I. and Førland, E. J. 2002b. Empirically downscaled temperature scenarios for Svalbard. *Atmos. Sci. Lett.* September 18, doi.10.1006/asle.2002.0051.

Biau, G., Zorita, E., von Storch, H. and Wackernagel, H. 1999. Estimation of precipitation by kriging in EOF space. *J. Climate* **12**, 1070–1085.

Bretherton, C. S., Smith, C. and Wallace, J. M. 1992. An intercomparison of methods for finding coupled patterns in climate data. *J. Climate* **5**, 541–560.

Busuioc, A., Chen, D. and Hellström, C. 2001. Performance of statistical downscaling models in GCM validation and regional climate change estimates: application for Swedish precipitation. *Int. J. Climatol.* **21**, 557–578.

Christensen, O. B., Christensen, J. H., Machenhauer, B. and Botzet, M. 1998. Very high-resolution climate simulations over Scandinavia—present climate. *J. Climate* **11**, 3204–3229.

Christensen, J. H., Räisänen, J., Iversen, T., Bjørge, D., Christensen, O. B. and Rummukainen, M. 2001. A synthesis of regional climate change simulations—a Scandinavian perspective. *Geophys. Res. Lett.* **28**(**6**), 1003.

Corte-Real, J., Qian, B. and Xu, H. 1998. Regional climate change in Portugal: precipitation variability associated with large-scale atmospheric circulation. *Int. J. Climatol.* **18**, 619–635.

Deliang, C. and Hellström, C. 1999. The influence of the North Atlantic Oscillation on the regional temperature variability in Sweden: spatial and temporal variations. *Tellus* **51A**, 505–516.

Ellner, S. P. 2001. Review of R, Version 1.1.1. *Bull. Ecol. Soc. Am.* **82**, 127–128.

Flury, B. 1988. *Common Principal Components and Related Multivariate Models, Wiley Series in Probability and Mathematical Statistics*. Wiley, New York.

Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. and Webster, M. D. 2002. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117.

Frich, P., Alexandersson, H., Ashcroft, J., Dahlström, B., Demarée, G. R., Drebs, A., van Engelen, A. F. V., Førland, E. J., Hanssen-Bauer, I., Heino, R., Jónsson, T., Jonasson, K., Nordli, P. Ø., Schmidth, T., Steffensen, P., Tuomenvirta, H. and Tveito, O. E. 1996. *North Atlantic Climatological Dataset (NACD Version 1)—Final Report*. Scientific Report 1. DMI, Copenhagen, Denmark.

Gebhardt, A. 2000. The SGEOSTAT package, v.1.0-10. URL: http://cran.r-project.org/. A geostatistical Package for R ("GNU S-plus") ported from S-plus (J.J. Majure).

Gentleman, R. and Ihaka, R. 2000. Lexical scope and statistical computing. *J. Computat. Graphical Statist.* **9**, 491–508.

Glahn, H. R. 1968. Canonical correlation and its relationship with discriminant analysis and multiple regression. *J. Atmos. Sci.* **25**, 23–31.

Grotch, S. and MacCracken, M. 1991. The use of general circulation models to predict regional climate change. *J. Climate* **4**, 286–303.

Haugen, J. E., Bjørge, D. and Nordeng, T. E. 1999. A 20-year climate change experiment with HIRHAM, using MPI boundary data. In: *RegClim. General Technical Report* (edsIversen, T. and Høiskar, B. A. K), 37–44, no. 3. NILU: http://www.nilu.no/regclim/.

Hellström, C., Chen, D., Achberger, C. and Räisanen, J. 2001. Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. *Climate Res.* **19**, 45–55.

Heyen, H., Zorita, E. and von Storch, H. 1996. Statistical downscaling of monthly mean North Atlantic air-pressure to sea level anomalies in the Baltic Sea. *Tellus* **48A**, 312–323.

Holland, M. M., Bitz, C. M. and Weaver, A. J. 2001. The influence of sea ice physics on simulations of climate change. *Geophys. Res. Lett.* **106**, 19,639–19,655.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K. and Johnson, C. A. 2001. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of IPCC*. International Panel on Climate Change (available from www.ipcc.ch).

Huth, R. 2002. Statistical downscaling of daily temperature in Central Europe. *J. Climate* **15**, 1731–1742.

Huth, R. and Kyselý, J. 2000. Constructing site-specific climate change scenarios on a monthly scale. *Theor. Appl. Climatol.* **66**, 13–27.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Wollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteoral Soc.* **77** (March), 437–471.

Kidson, J. W. and Thompson, C. S. 1998. A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate* **11**, 735–753.

Kilsby, C. G., Cowpertwait, P. P. P., O'Connel, P. E. and Jones, P. D. 1998. Predicting rainfall statistics in England and Wales using atmospheric circulation variables. *Int. J. Climate* **18**, 523–539.

Kutzbach, J. 1967. Empirical eigenvectors of sea-level pressure, surface temperature, and precipitation complexes over North America. *J. Appl. Meterol.* **6**, 791–802.

Lorenz, E. N. 1956. *Empirical Orthogonal Functions and Statistical Weather Prediction*. Scientific Report 1. Department of Meteorology, MIT, USA.

Matheron, G. 1963. Principles of geostatistics. *Economic Geol.* **58**, 1246–1266.

Murphy, J. 1999. An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate* **12**, 2256–2284.

Murphy, J. 2000. Predictions of climate change over Europe using statistical and dynamical downscaling techniques. *Int. J. Climatol.* **20**, 489–501.

North, G. R., Bell, T. L. and Cahalan, R. F. 1982. Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.* **110**, 699–706.

Oshima, N., Kato, H. and Kadokura, S. 2002. An application of statistical downscaling to estimate surface air temperature in Japan. *J. Geophys. Res.* **107**(D10), ACL–14.

Palmer, T. N. and Räisänen, J. 2002. Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature* **415**, 512–514.

Palmer, T. N., Branckovic, C. and Richardson, D. S. 2000. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Met. Soc.* **126**, 2013–2034.

Preisendorfer, R. W. 1988. *Principal Component Analysis in Meteorology and Oceanology*. Elsevier, Amsterdam.

Räisänen, J. 2001a. CO_2-induced climate change in CMIP2 experiments: quantification of agreement and role of internal variability. *J. Climate* **14**, 2088–2104.

Räisänen, J. 2001b. Intercomparison of 19 global climate change simulations from an Arctic perspective. In: *Report from the Arctic Climate Impact Assessment Modeling and Scenario Workshop*. (edsKällen, E., Kattsov, V., Walsh, J. and Weatherhead, E.), 11–13. ACIA Secretariat, PO Box 757740, Fairbanks, AK 99775-7740 (http://www.acia.uaf.edu).

Räisänen, J. 2002. CO_2-induced changes in interannual temperature and precipitation variability in 19 CMIP experiments. *J. Climate* **15**, 2395–2411.

Räisänen, J. and Alexandersson, H. 2003. A probabilistic view on recent and near future climate change in Sweden. *Tellus* **55A**, 113–125.

Räisänen, J. and Palmer, T. N. 2001. A probability and decision-model analysis of a multimodel ensemble of climate change simulations. *J. Climate* **14**, 3212–3226.

Rigor, I. G., Wallace, J. M. and Colony, R. L. 2002. Response of Sea Ice to the Arctic Oscillation. *J. Climate* **15**, 2648–2663.

Schubert, S. 1998. Downscaling local extreme temperature change in south-eastern Australia from the CSIRO MARK2 GCM. *Int. J. Climatol.* **18**, 1419–1438.

Sengupta, S. K. and Boyle, J. S. 1993. *Statistical Intercomparison of Global Climate Models: a Common Principal Component Approach*. Technical Report 13. PCMDI, Lawrence Livermore National Laboratory, California, USA (http://www-pcmdi.llnl.gov/pcmdi/pubs/pdf/13.pdf).

Sengupta, S. and Boyle, J. S. 1998. Using common principal components in comparing GCM simulations. *J. Climate* **11**, 816–830.

Stott, P. A., Tett, S. F. B., Jones, G. S., Allen, M. R., Ingram, W. J. and Mitchell, J. F. B. 2001. Attribution of twentieth century temperature change to natural and anthropogenic causes. *Climate Dynam.* **17**, 1–21.

Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. and Mitchell, J. F. B. 1999. Causes of twentieth-century temperature change near the Earth's surface. *Nature* **399**, 569–572.

von Storch, H., Zorita, E. and Cubasch, U. 1993a. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *J. Climate* **6**, 1161–1171.

von Storch, H., Zorita, E. and Cubasch, U. 1993b. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *J. Climate* **6**, 1161–1171.

Wilby, R. L. 1997. Non-stationarity in daily precipitation series: Implications for GCM down-scaling using atmospheric circulation indices. *Int. J. Climate* **17**, 439–454.

Wilby, R. L., Hassan, H. and Hanaki, K. 1998. Statistical downscaling of hydrometeorological variables using general circulation model output. *J. Hydrology* **205**, 1–19.

Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, Orlando, FL.

Zorita, E. and von Storch, H. 1997. *A Survey of Statistical Downscaling Results*. Technical Report 97/E/20. GKSS.

Zorita, E. and von Storch, H. 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J. Climate* **12**, 2474–2489.